

Методы, основанные на предположении нормальности, и близкие к ним

- Линейный и квадратичный дискриминант
- Дискриминант Фишера
- Логистическая регрессия.
- Случай независимых переменных.

Нормальное распределение

Одномерный случай.

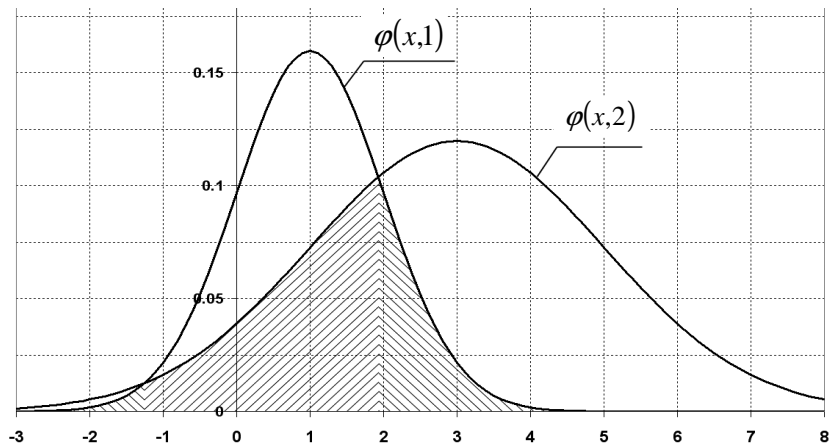
Заданы $P(y)$ и условные плотности вероятности $\varphi_y(x)$,
 $y \in \{1, 2\}$

$$\varphi_y(x) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(x-\mu_y)^2}{2(\sigma_y)^2}}.$$

Совместная плотность вероятности

$$\varphi(x, y) = \varphi_1(x)P(1) + \varphi_2(x)P(2).$$

Совместная плотность вероятности



Условные вероятности

Функции условной вероятности:

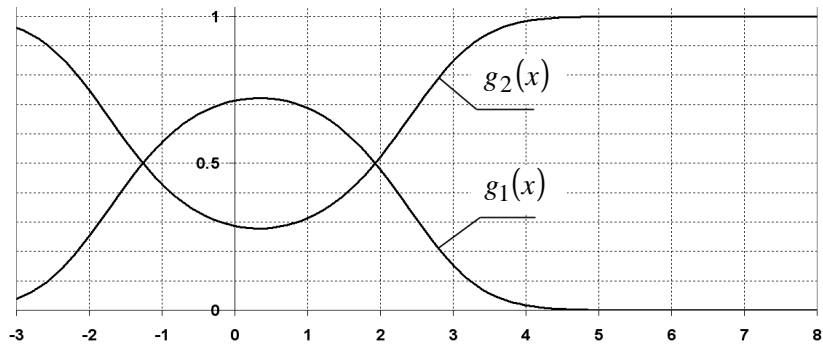
$$g_1(x) = P(y = 1/x) = \frac{\varphi(x,1)}{\varphi(x)} = \frac{\varphi(x,1)}{\varphi(x,1) + \varphi(x,2)},$$

$$g_2(x) = P(y = 2/x) = 1 - g_1(x).$$

Байесовская решающая функция

$f(x) = 1$, при $g_1(x) \geq g_2(x)$; $f(x) = 2$, при $g_1(x) < g_2(x)$.

Условные вероятности



Разделяющая функция

Байесовская решающая функция

$$f(x) = \begin{cases} 1, & l(x) \geq 0 \\ 2, & l(x) < 0 \end{cases}.$$

Разделяющая функция

$$l(x) = \ln \frac{g_1(x)}{g_2(x)} = \ln \varphi_1(x) - \ln \varphi_2(x) + \ln \frac{P(1)}{P(2)}.$$

Для нормальных плотностей

$$l(x) = \frac{(x - \mu_2)^2}{(\sigma_2)^2} - \frac{(x - \mu_1)^2}{(\sigma_1)^2} + \ln \frac{\sigma_2}{\sigma_1} + \ln \frac{P(1)}{P(2)}.$$

Многомерный нормальный закон

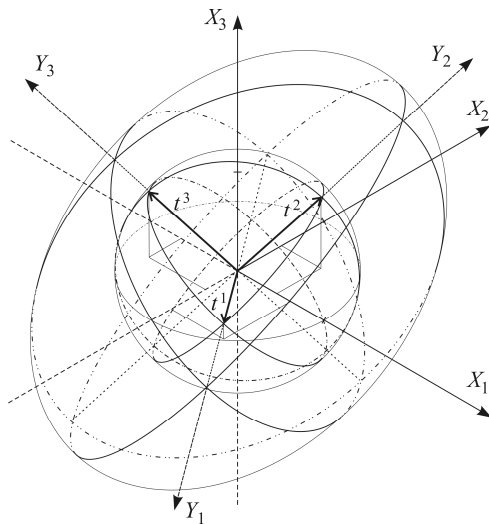
Плотность вероятности

$$\varphi_k = \frac{1}{(2\pi)^{n/2} |\lambda_k|^{n/2}} e^{-\frac{1}{2} Q_k(x)},$$

где $Q_k(x) = (x - \mu_k)' (\lambda_k)^{-1} (x - \mu_k)$,

μ_y — вектор средних для класса y ,

λ_y — ковариационная матрица.



эллипсоид
рассеяния

Разделяющая функция

Подставляем нормальные плотности

$$2l(x) = Q_2(x) - Q_1(x) + \ln|\lambda_2| - \ln|\lambda_1| + 2\ln P_1 - 2\ln P_2.$$

После преобразований

$$2l(x) = xAx + bx + c, \text{ где}$$

$$A = (\lambda_2)^{-1} - (\lambda_1)^{-1}, \quad b = 2\mu_1(\lambda_1)^{-1} - 2\mu_2(\lambda_2)^{-1},$$

$$c = \mu_2(\lambda_2)^{-1}\mu_2 - \mu_1(\lambda_1)^{-1}\mu_1 + \ln|\lambda_2| - \ln|\lambda_1| + 2\ln P_1 - 2\ln P_2$$

Оценивание параметров

Априорные вероятности классов: $\tilde{P}_k = \frac{N_k}{N}$.

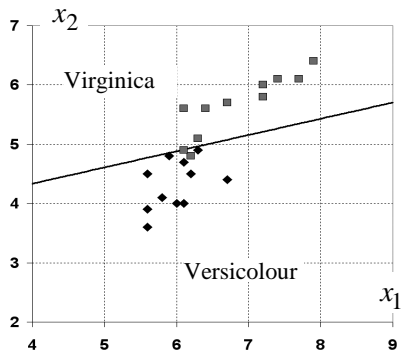
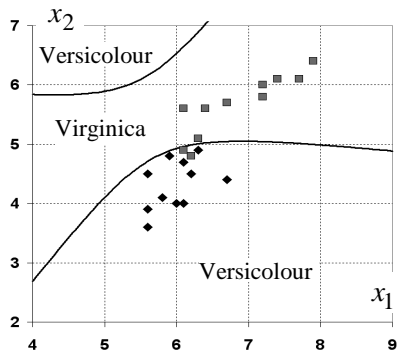
Компоненты вектора средних

$$\tilde{\mu}_j^k = \frac{1}{N_k} \sum_{i \in I_k} x_{ij} \ .$$

Оценка ковариационной матрицы

$$\tilde{\lambda}_{jl}^k = \frac{1}{N_k} \sum_{i \in I_k} (x_{ij} - \tilde{\mu}_j^k)(x_{il} - \tilde{\mu}_l^k) \ .$$

Задача Iris



Замечания

В реальных задачах довольно редко можно обосновать нормальность распределений.

Даже если известно, что распределения действительно нормальны, это не гарантирует оптимальность выборочной решающей функции.

При большом числе переменных и малой выборке приходится строить решение как при равных матрицах ковариаций.

Дискриминант Фишера

Идея заключается в выборе направления, при проецировании выборки на которое образы классов оказываются наиболее удалёнными друг от друга

$$\Phi(w) = \frac{(\tilde{\mu}_{w,1} - \tilde{\mu}_{w,-1})^2}{\tilde{S}_{w,1} + \tilde{S}_{w,-1}},$$

где $\tilde{\mu}_{w,y} = \frac{1}{N_y} \sum_{i=1}^N w x^i \cdot I(y^i = y)$ — среднее,

а $\tilde{S}_{w,y} = \frac{1}{N_y} \sum_{i=1}^N (w x^i - \tilde{\mu}_{w,y})^2 \cdot I(y^i = y)$ — средний квадрат отклонений проекций.

Критерий приводится к виду

$$\Phi(w) = \frac{(w\tilde{\mu}_1 - w\tilde{\mu}_{-1})^2}{w^T(\tilde{S}_1 + \tilde{S}_{-1})w} = \frac{w^T(\tilde{\mu}_1 - \tilde{\mu}_{-1})(\tilde{\mu}_1 - \tilde{\mu}_{-1})^T w}{w^T \tilde{S} w},$$

где $\tilde{\mu}_y = \frac{1}{N_y} \sum_{i=1}^N x^i \cdot I(y^i = y)$ – среднее точек выборки y -го класса, \tilde{S}_y – выборочная ковариационная матрица y -го класса, $\tilde{S} = \tilde{S}_1 + \tilde{S}_{-1}$.

Последняя форма критерия имеет вид отношения Релея.

Известно, что максимум $\Phi(w)$ достигается при

$$w = w_\phi = \tilde{S}^{-1}(\tilde{\mu}_1 - \tilde{\mu}_{-1}).$$

Замечания

Выражение для w_Φ очень похоже на выражение для нормали к разделяющей гиперплоскости для случая нормальных распределений с равными матрицами ковариаций. Отличие лишь в том, что во втором случае вместо \tilde{S} используется $\frac{1}{N}(N_1\tilde{S}_1 + N_{-1}\tilde{S}_{-1})$.

Метод предполагает оценивание по выборке только n параметров и не требователен к объёму выборки.

Логистическая регрессия

Имеем

$$g_1(x) = P(y=1/x) = \frac{P\varphi_1(x)}{P(1)\varphi_1(x) + (1-P(1))\cdot\varphi_{-1}(x)}.$$

Подставим нормальную плотность

$$g(x) = \frac{1}{1 + e^{-(wx+w_0)}} = \sigma(wx + w_0),$$

где $w = \frac{1}{2}S^{-1}(\mu_1 - \mu_{-1})$,

$$w_0 = \frac{1}{2}\mu_{-1}^T S^{-1}\mu_{-1} - \frac{1}{2}\mu_1^T S^{-1}\mu_1 + \ln P(1) - \ln P(2).$$

Сигмоид

$\sigma(z) = \frac{1}{1+e^{-z}}$ – так называемая логистическая функция

(иногда также называемая сигмоидом или логит-функцией).

Метод логистической регрессии основан на оценивании функции условной вероятности моделью $\tilde{g}(x) = \sigma(\tilde{w}x + \tilde{w}_0)$, в которой \tilde{w} и \tilde{w}_0 – настраиваемые параметры.

На практике параметры модели обычно оцениваются путём максимизации критерия правдоподобия (условной вероятности выборки).

Случай независимых переменных

Из формулы Байеса можем записать

$$\begin{aligned} g(x) &= P(y = 1/x) = \\ &= \frac{P(dx, y = 1)}{P(dx, y = 1) + P(dx, y = -1)} = \frac{1}{1 + \frac{1-p}{p} \cdot \frac{P(dx/y = -1)}{P(dx/y = 1)}}. \end{aligned}$$

$$\text{Пусть } P(dx/y) = \prod_{j=1}^n P(dx_j/y).$$

Подставив это произведение в предыдущее выражение, после преобразований имеем

$$\frac{p}{1-p} \cdot \left(\frac{1}{g(x)} - 1 \right) = \prod_{j=1}^n \frac{p}{1-p} \cdot \left(\frac{1}{g_j(x_j)} - 1 \right),$$

$$\text{где } g_j(x_j) = P(y=1/x_j) = \frac{P(dx_j, y=1)}{P(dx_j)}.$$

Логарифмируем последнее выражение и получаем

$$\sigma^{-1}(g(x)) = (n-1)(\ln p - \ln(1-p)) + \sum_{j=1}^n \sigma^{-1}(g_j(x_j)),$$

где $\sigma^{-1}(u) = \ln\left(\frac{1}{u} - 1\right)$ – функция, обратная сигмоиду

$$\sigma(z) = \frac{1}{1+e^{-z}}.$$

Заметим, что полученное выражение имеет вид логистической регрессии, а именно

$$g(x) = \sigma \left(w_0 + \sum_{j=1}^n w_j \sigma^{-1}(g_j(x_j)) \right),$$

при $w_0 = (n-1)(\ln p - \ln(1-p))$, $w_j = 1$.

Границы применимости вероятностной постановки

Данная задача известна как «парадокс конвертов».

Игроку предлагается выбрать один из двух одинаковых на вид запечатанных конвертов с деньгами, причём известно, что сумма в одном из них в 10 раз больше, чем в другом. При этом игроку разрешается вскрыть один конверт, после чего решить, забрать его или оставшийся запечатанным.

Предположим, что игрок открыл наугад взятый конверт и обнаружил в нём 10 долларов. Получается, что в другом конверте можно в равной степени ожидать сумму в 1 доллар или в 100 долларов. В среднем получаем 50 с половиной.

Следует ли из этих рассуждений, что игроку выгоднее отказаться от первоначально выбранного конверта и забрать себе другой конверт? Можно ли считать, что математическое ожидание выигрыша в этом случае (для рассмотренного примера) составит 50,5 долларов?